

On the Equivalence of Restricted Boltzmann Machines and Tensor Network States

Jing Chen,¹ Song Cheng,¹ Haidong Xie,¹ Lei Wang,^{1,*} and Tao Xiang^{1,2,†}

¹*Institute of Physics, Chinese Academy of Sciences, P.O. Box 603, Beijing 100190, China*

²*Collaborative Innovation Center of Quantum Matter, Beijing 100190, China*

Restricted Boltzmann machine (RBM) is one of the fundamental building blocks of deep learning. RBM finds wide applications in dimensional reduction, feature extraction, and recommender systems via modeling the probability distributions of a variety of input data including natural images, speech signals, and customer ratings, etc. We build a bridge between RBM and tensor network states (TNS) widely used in quantum many-body physics research. We devise efficient algorithms to translate an RBM into the commonly used TNS. Conversely, we give sufficient and necessary conditions to determine whether a TNS can be transformed into an RBM of given architectures. Revealing these general and constructive connections can cross-fertilize both deep learning and quantum-many body physics. Notably, by exploiting the entanglement entropy bound of TNS, we can rigorously quantify the expressive power of RBM on complex datasets. Insights into TNS and its entanglement capacity can guide the design of more powerful deep learning architectures. On the other hand, RBM can represent quantum many-body states with fewer parameters compared to TNS, which may allow more efficient classical simulations.

I. INTRODUCTION

Deep learning is transforming the world with its far-reaching applications in computer vision, speech recognition, natural language processing, recommender systems, etc [1, 2]. At the core of many of these applications are the artificial neural networks which can be used to recognize or even discover interesting patterns in the input data [3–5]. In a nutshell, the neural nets act as trainable functional mappings of multiple variables. To design even more powerful and intelligent machines requires one to quantify and extend the expressive power of the neural nets. However, there is a gap between the mathematical foundation and the real-world applications which are largely driven by the engineering practices [2], because it has long been a difficult endeavor to rigorously quantify the neural nets over complicated datasets.

Insights into the physical rules governing the neural networks may offer an answer to the great success of deep learning and guide its more fruitful development in the future. For example, statistical physics has a long-standing impact on machine learning [4, 6], because both fields concern about collective behavior emerged from a large amount of microscopic degree of freedoms. Moreover, as suggested by Mehta and Schwab [7], there is a connection between the deep learning and the renormalization group. Lin and Tegmark [8] also argued that the “unreasonable success” of deep learning can be traced back to the law of physics, which often imposes symmetry, locality, compositionality, polynomial log-probability, and other properties on the input data.

RBM is a vivid example of the intrinsic connection between statistical physics and machine learning. An RBM is a special type of neural networks which can be better

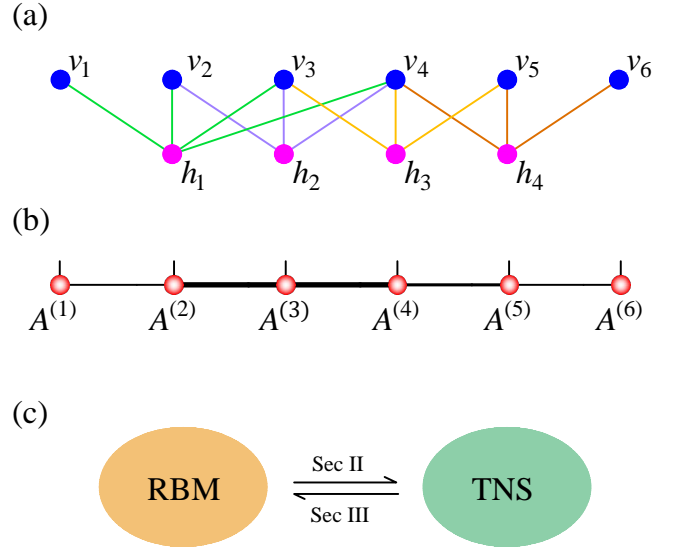


Figure 1. (a) The RBM function defined in Eq. (2). The blue dots represent visible units v and the magenta dots represent the hidden units h . They are coupled through the connections indicated by the solid lines. (b) The MPS defined in Eq. (3). Each red dot denotes a three index tensor $A^{(i)}$. Throughout this paper, we use flat dots to represent units of an RBM, and use stereoscopic dots to represent tensors. (c) Our paper builds explicit connections between RBM and TNS. See Sec. II and Sec. III for detailed discussions on the mapping between them.

understood as an energy-based model. As illustrated in Fig. 1(a), it consists of a set of interconnected visible and hidden binary variables. These variables are assumed to satisfy the Boltzmann distribution whose energy functional is defined by

$$E(v, h) = - \sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i W_{ij} h_j, \quad (1)$$

where $v = \{v_i\}$ and $h = \{h_j\}$ are the visible and hidden

* wanglei@iphy.ac.cn

† txiang@iphy.ac.cn

binary variables. We denote their number as n_v and n_h respectively. Parameters a_i, b_j are the biases applied to the visible and hidden units, respectively. W_{ij} is the coupling matrix between these two units.

By integrating out the hidden units, the RBM represents the marginal distribution of the visible variables (omitting an irrelevant normalization factor)

$$\begin{aligned}\Psi_{\text{RBM}}(v) &= \sum_h e^{-E(v,h)} \\ &= \prod_{i,j} e^{a_i v_i} \left(1 + e^{b_j + \sum_i v_i W_{ij}} \right),\end{aligned}\quad (2)$$

In the RBM there is no direct connection between the visible units. However the hidden units generate effective connections or interactions among them. By increasing the number of hidden units and connections, the RBM can in principle parameterize more complex functions of the visible units [9–11].

One can train an RBM by specifying its parameters such that the probability distribution of the visible units reproduces that of the input data [12, 13]. The hidden units of a trained RBM may also reveal correlations in the data with appealing physical meanings. For example, in an RBM trained with a dataset of images containing handwritten digits, the connection weight contains the information about the pen strokes [14]. These learning features can be used either for discriminative tasks, such as pattern recognition, or for generative tasks, such as generating more samples according to the learned distribution. RBM has played an important role in the recent renaissance of deep learning [15, 16], because of its versatile abilities in feature extraction and dimensionality reduction of complex data sets.

Recently, RBM has attracted great attention in the field of quantum many-body physics. For example, Carleo and Troyer [17] proposed an RBM inspired variational wavefunction to study quantum many-body systems at or away from equilibrium. Deng *et al.* [18] constructed exact RBM representations for several interesting topological states. Torlai and Melko [19] trained an RBM to reproduce the thermodynamics of a statistical physics model. Huang and Wang [20] used RBM as a recommender system to accelerate Monte Carlo simulation of quantum many-body systems. Liu *et al.* [21] reported similar ideas using classical spin models instead of the RBM.

These developments have raised several critical questions about the expressive power of neural nets. Is RBM more expressive than the standard variational wavefunctions of quantum states [17]? Can RBM efficiently describe the probability distribution of physical models at criticality [19, 20]? Unfortunately, the existing universal approximation theorem [9–11] established in the machine learning community is not particularly instructive because it involves exponentially large resources, and it cannot be used as a guiding principle to solve physical or industrial problems.

In fact, the quest for more expressive wavefunction is actually central to the quantum many-body physics research. Ideally, those ansatz should accurately describe exponentially large degree of freedoms with polynomial resources. Tensor network states (TNS) [22] are promising candidates to meet this demand. Fig. 1(b) shows one of the simplest TNS, the matrix product state (MPS) [23], as an example. The MPS parametrizes a wavefunction of n_v physical variables as,

$$\Psi_{\text{MPS}}(v) = \text{Tr} \left(\prod_{i=1}^{n_v} A^{(i)}[v_i] \right), \quad (3)$$

where each $A^{(i)}$ is a 3-index tensor indicated by a red dot in Fig. 1(b). For fixed physical variables on the vertical lines, each $A^{(i)}[v_i]$ is a matrix. We denote their matrix sizes as the virtual bond dimensions indicated by the thickness of the horizontal bonds in Fig. 1(b). Connecting these horizontal bonds means tensor contraction of the virtual degree of freedoms. By increasing the bond dimension of the virtual bonds, MPS could represent increasingly complex multivariable functions [23]. MPS representation is equivalent to the tensor train decompositions in the applied math community [24]. Similarly, one can connect higher order tensors in a two dimensional network to represent the physical degree of freedoms lies in two dimensions more naturally. This construction is named projected entangled pair state (PEPS) [25]. In the past decades, quantum physicists have developed solid theoretical understanding and efficient numerical techniques for TNS. See [22, 26] for pedagogical reviews on TNS.

The physics community also has an answer for TNS’s “unreasonable effectiveness”: the entanglement area law [27]. Entanglement entropy [28] characterizes the information content of a bipartition of the system. And the entanglement area law means that the entanglement entropy increases only proportional to the boundary between the two subsystems. Although there are extremely large number of possible quantum states, many physical states of practical interests fulfill the entanglement area law [27]. TNS are designed to efficiently represent these quantum states with relatively low entanglement entropy and have achieved remarkable successes in the past decades [29].

The success of TNS for quantum many-body systems makes one wonder whether is there anything similar to quantify the expressive power of deep learning architectures. As Refs. [2, 8] suggest, we only care about a tiny fraction of the input data among infinite number of possible inputs in the practical machine learning applications. This hints for certain guiding principle similar to the entanglement area law to quantify the required resources.

In this paper, we present a general and constructive connection between RBM and TNS. With this correspondence, many concepts and techniques from deep learning and quantum physics can be exchanged. By transforming an RBM to a TNS and exploiting its entanglement en-

tropy bound, one can quantify the expressibility of RBM for quantum states, for statistical physics models, and for industrial datasets. On the other hand, we present sufficient and necessary conditions on TNS so that it can be transformed into an RBM of a given structure. RBM can serve as a very efficient classical representation of the quantum states.

The organization of this paper is as follows: In Sec. II, we present algorithms to transform the RBM into TNS such as MPS and PEPS and discuss their consequences. In Sec. III, we discuss the sufficient and necessary conditions of a TNS to have an RBM representations of a given architecture. In Sec. IV, we illustrate the RBM-TNS connection by showing the four ground states of the toric code model in the RBM form and in the corresponding TNS form. In Sec. V, we discuss the redundancy of RBM parametrization and illustrated how to simplify them using TNS methods. In Sec. VI, we discuss implications of our results both for machine learning and for quantum many-body physics. In Appendix A we show the equivalence of the shift-invariant RBM employed in Ref. [17] to an MPS and discuss their entanglement entropy bounds. In Appendix B, we give a sufficient condition to find the RBM representation of some specific TNS for the statistical physics models and simple quantum states. Finally, in Appendix C we discuss the general equivalence between more general Boltzmann machines [30, 31] and TNS.

II. TNS REPRESENTATION OF RBM: THE ROUTE TO MPS AND PEPS

In this section we show the general equivalence of an RBM and an MPS via a constructive approach. Namely, given the RBM architecture and parameters $\{W_{ij}, a_i, b_j\}$, we present algorithms to obtain the tensors explicitly. We first present a simple and intuitive approach, then a more sophisticated approach which provides a tighter bound on the tensor bond dimension. For both algorithms we provide software implementations in the supplementary materials [32].

Before proceeding, we first clarify a few notations about the RBM. Firstly, in the standard machine learning applications and in [19, 20], one tunes the real parameters $\{W_{ij}, a_i, b_j\}$ to let the RBM [Eq. (2)] model the probability distributions of the input data. While references [17, 18] interpret Eq. (2) as the amplitude of quantum mechanical wavefunctions. Therefore, these authors generalized the weights and biases to complex numbers. In this paper, we consider complex RBM parameters for generality.¹ Secondly, conventionally one views the RBM as an energy based generative model. While for our discussion about the expressibility of Eq. (2) it is sufficient

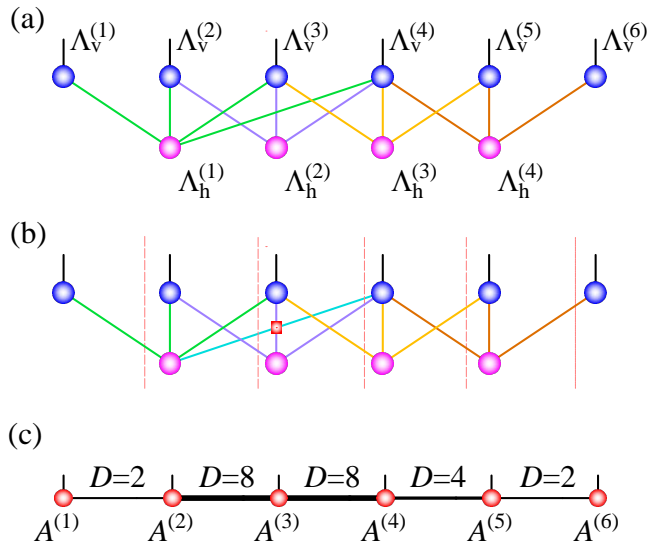


Figure 2. Transform an RBM to an MPS. (a) A TNS with the same network structure as the RBM in Fig. 1(b). The blue dots represent the diagonal tensors $\Lambda_v^{(i)}$ at the visible units and the magenta dots represent the $\Lambda_h^{(j)}$ at the hidden units, see Eqs. (5,6). The bonds denote the matrix $M^{(ij)}$ defined in Eq. (4). (b) The TNS is cut into n_v pieces. The long range connection is mediated by a cross gate (red square) attached to the vertical connection. (c) The MPS representation. The 3 index tensor $A^{(i)}$ (red dot) is determined by contraction of the internal indices of each piece in (b). The bond dimension D of the MPS is determined by the number of cut connections of the RBM.

to view it as a function approximator such as a feedforward neural net [20, 33]. Thirdly, although the standard RBM has dense all-to-all connections between the visible and the hidden units, for generality and clarity we illustrated the transformation using the RBM with sparser connections shown in Fig. 1(a). Our results are nevertheless general and apply to the RBM with dense inter connections between the visible and the hidden units.

To start with, we rewrite the RBM in terms of a TNS with external (visible) and internal (hidden) degree of freedoms while keeping the network structure unchanged, see Fig. 2(a). We define a diagonal tensor $\Lambda_v^{(i)}$ and $\Lambda_h^{(j)}$ on each visible and hidden unit indicated by the dots. And we introduce a 2×2 matrix $M^{(ij)}$ on each connection line between the visible unit v_i and the hidden unit h_j . The tensor elements are related to the weights and biases of the RBM

$$M^{(ij)} = \begin{pmatrix} 1 & 1 \\ 1 & e^{W_{ij}} \end{pmatrix}, \quad (4)$$

$$\Lambda_v^{(i)} = \text{diag}(1, e^{a_i}), \quad (5)$$

$$\Lambda_h^{(j)} = \text{diag}(1, e^{b_j}), \quad (6)$$

where $\text{diag}(\cdot, \cdot)$ denotes a diagonal tensor with nonzero elements given in the bracket.

To translate the TNS in Fig. 2(a) into an MPS, we

¹ The references [17, 18] defines RBM with units with ± 1 instead of binary values. This amounts to a simple rescaling and offset of the weights and biases.

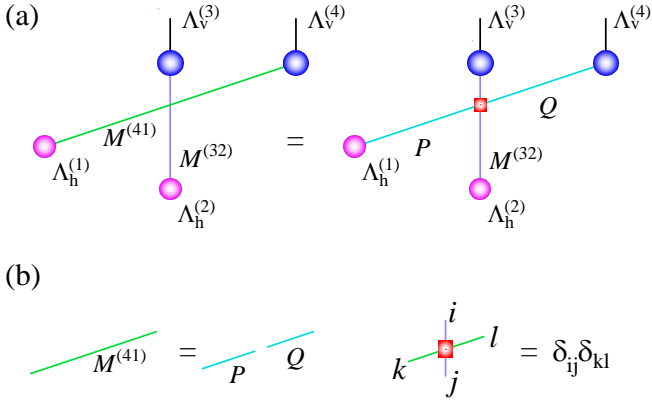


Figure 3. (a) The cross gate indicated by the red square mediates the long range connections between v_4 and h_1 . The green connection denotes the matrix $M^{(41)}$ defined in Eq. (4). The cyan connections denote matrices P, Q which double the virtual bond dimensions. (b) The cross gate splits the matrix $M^{(41)}$ into the product of two arbitrary matrices P and Q .

next divide it into n_v pieces as shown in Fig. 2(b). Contracting all the internal indices within each piece is then equivalent to summing over the hidden units. Finally, we merge the external connections between pieces into the virtual bonds of the MPS shown in Fig. 2(c). The thickness of the virtual bonds indicates the corresponding bond dimension of the MPS.

From the step Fig. 2(b) to (c), we need to pay special attention to the “long-range” connections indicated by the cyan line which crosses the in between pieces. We introduce a “cross gate” denoted by the red square to break this long range connection into segments. As explained in Fig. 3, the cross gate splits the matrix $M^{(41)}$ into the product of two arbitrary matrices $M^{(41)} = P \cdot Q$. Afterwards, the matrices P and Q are absorbed into $A^{(2)}$ and $A^{(3)}$ respectively. The “long-range” connection doubles the bond dimension of $A^{(3)}$ to $D = 8$. In general, long range connections will double the bond dimension of all tensors they pass by. Following the scheme of Fig. 2 one can obtain an MPS representation of arbitrary RBM. The MPS bond dimensions are given by the number of connections n cut at the bipartition $D = 2^n$.

The mapping from the RBM to the MPS shown in Fig. 2 is not unique because one can assign the hidden units into different subregions. The resulting MPS will have different bond dimensions even though they represent the same function. One can bring these equivalent MPSs into a unique canonical form [34, 35] by performing the Schmidt decomposition for each bond. See Sec. V for more discussions about the redundancies in the RBM and TNS representations.

The explicit connection between the RBM and the MPS allows one to quantify the RBM’s expressibility using the entanglement entropy because it is convenient to read out the entanglement information in TNS language. We divide the visible units into two sets $v = \{v_{i \in A}, v_{i \in B}\}$

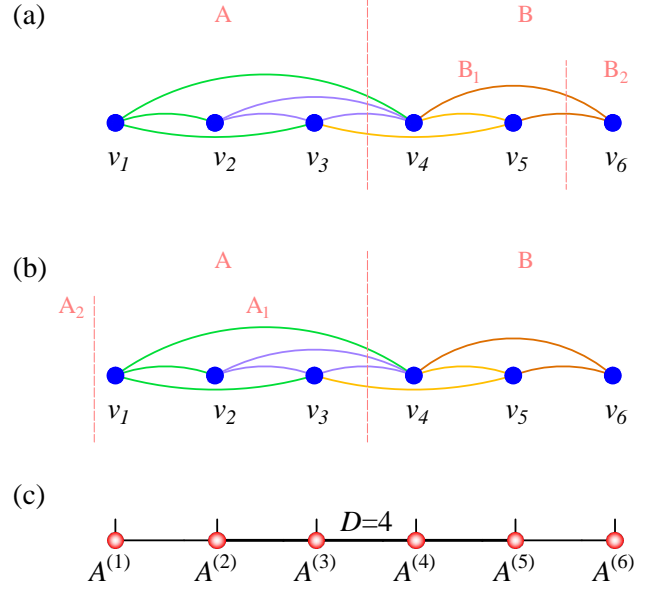


Figure 4. Estimating a tighter entanglement entropy bound. (a) We show only the visible units of the RBM and the connections mediated by the hidden units as the arcs. A division of the system into A and $B = B_1 \cup B_2$, where B_1 is the interface region can be reached from A . (b) A division of the system into $A = A_1 \cup A_2$ and B , where A_1 is the interface region can be reached from B . (c) The resulting MPS has a smaller bond dimension at the bipartition compared to Fig. 2(c).

and define the reduced density matrix

$$\rho = \sum_{\{v_{i \in B}\}} \Psi_{\text{RBM}}^* (\{v'_{i \in A}, v_{i \in B}\}) \Psi_{\text{RBM}} (\{v_{i \in A}, v_{i \in B}\}). \quad (7)$$

The von Neumann entanglement entropy is defined as $S = -\text{Tr}(\rho \ln \rho)$ [28]. The entanglement entropy characterizes the information content of the distribution Ψ_{RBM} . The entanglement entropy can be viewed as a proxy for the correlations between the two partitions. Uncorrelated data has zero entanglement entropy at any bipartition. While long range correlations typically increases the entanglement. We conjecture that the realistic dataset in most successful machine learning applications has relatively low entanglement entropies. Computationally, the entanglement entropy of the MPS at each bipartition can be obtained as a byproduct of the canonicalization. The maximal entanglement entropy is bounded by the MPS bond dimension $\ln D$, in which case all the eigenvalues of the reduced density matrix are $1/D$ [36].

The bond dimensions obtained by the simple and intuitive approach Fig. 2 is only an upper bound. For example, the bond dimension of the second bond $D = 8$ in Fig. 2(c) is more than enough to capture the entanglement since there are only two sites on the left. Next we present a more sophisticated approach to provide a tighter bound on the bond dimensions than Fig. 2(c). This approach is also independent of the assignment of the hidden units. Figure 4(a) shows an RBM with the

hidden units been traced out, where the arcs indicate the hidden units mediated connection between the visible units. The bipartition of the RBM cut the visible units into two parts A and $B = B_1 \cup B_2$, where B_1 contains the visible units can be reached directly from A , and B_2 contains the remaining part. The RBM function Eq. (2) has a product form

$$\Psi_{\text{RBM}}(v) = \psi(\{v_{i \in A}, v_{i \in B_1}\}) \phi(\{v_{i \in B_1}, v_{i \in B_2}\}). \quad (8)$$

Once the visible units in B_1 are given the function reduced to a product state of $v_{i \in A}$ and $v_{i \in B_2}$. The tensor rank [37, 38] of Eq. (8) is determined by the visible degree of freedoms of B_1 .

Alternatively, one can also consider a subdivision of $A = A_1 \cup A_2$ depending on whether it can be reached directly from the units in the part B . In the case of Fig. 4(b), $A_1 = A$ and $A_2 = \emptyset$. The MPS bond dimension at the A, B bipartition is determined by the smaller one of the two interface regions $m = \min(|A_1|, |B_1|)$. When mapping to an MPS, the physical state of A_1 or B_1 is copied to the virtual bond. Compared to Fig. 2, the approach of Fig. 4 always results into an MPS with smaller bond dimensions since $m \leq n$. For example, the connections out of the unit v_4 are counted independently in Fig. 2(a), but some of them end into the same visible units. As a consequence, the bond dimension of the center bond in Fig. 4(c) is $D = 4$ instead of $D = 8$ in Fig. 2(c).

In general, if there are m units in the interface region of a bipartition of the RBM, the corresponding MPS bond dimension is

$$D = 2^m. \quad (9)$$

The MPS bond dimension of the Fig. 4(c) is $D = 2, 4, 4, 4, 2$ respectively, which is smaller than Fig. 2(c). The bond dimension determines the maximal entanglement entropy which can be captured by the RBM at the bipartition

$$S_{\text{max}} = m \ln 2. \quad (10)$$

The entanglement entropy bound Eq. (10) only depends on the connectivity structure of the RBM. Employing this central result one can quantify the expressive power of various RBM for data such as the natural images or the quantum many-body states [17, 18]. Note that Eq. (10) is only an upper bound of the entanglement entropy, to obtain the precise value one needs to construct the MPS representation explicitly. We refer the interested readers to the code implementation in the supplementary materials [32].

It is also instructive to compare the number of parameters of the RBM and the corresponding TNS parametrization of the same function. Assuming each visible unit has m effectively connections to the surrounding visible units, the corresponding tensor will have 2^m parameters on its virtual bonds. Thus, in principle the RBM can be exponentially more compact than the

TNS representation. On the other hand, one should also note that although the TNS representation has a nominal larger number of parameters, they may still be advantageous in practice for solving quantum many-body physics problems thanks to more efficient optimization methods [22].

To conform with the effective long range connections in Fig. 4, one can first arrange the visible units into a two dimensional lattice, then map it into a PEPS [25]. This is particularly relevant if the original RBM architecture already accommodate for the dimensionality of the data, for example, pixels of images or physical degree of freedoms reside in two dimension.

In general, suppose the interface region size m does not scale with the system size. The width of the region is in the order of $m^{\frac{1}{d}}$. The entanglement entropy bound scales as

$$S_{\text{max}} \sim m^{\frac{1}{d}} L^{d-1}, \quad (11)$$

where d is the spatial dimension, L is the linear size of the subregion. Thus, the maximum entanglement of a sparsely connected RBM [18] still follows the area law, similar to the conventional TNS such as MPS or PEPS. While for densely connected RBM each visible unit can reach all other visible units through the hidden units, thus $m = \min(|A|, |B|)$ for any bipartition. The entanglement entropy bound can therefore reach the volume law scaling. The dense RBM provides a compact representation of highly entangled quantum states. The number of parameters of the dense RBM scales polynomially with the system size. On the other hand, to describe the state with same entanglement one needs exponentially large number of parameters in the MPS or PEPS, or needs to use the TNS with hierarchical structures [39, 40]. The RBM-TNS correspondence justifies the excellent accuracy achieved in the variational study [17]. Appendix A discusses about the TNS representation and entanglement capacity of the shift-invariant RBM state employed in Ref. [17].

In fact, the discovered RBM-TNS correspondence is not limited by the “restricted” feature (i.e. no coupling within the visible or the hidden units) of the RBM. In Appendix C, we proved a general TNS can be always mapped to a Boltzmann machine [30]. Moreover, the mappings shown in Fig. 2 and Fig. 4 apply as well to the deep Boltzmann machines [31] with multilayers of the hidden units. The visible units in a deep Boltzmann machine can have longer ranged effective connections mediated by the deep hidden units, therefore achieve a larger entanglement capacity compared to the RBM with the same number of hidden units and connections. Thus, the entanglement entropy bound Eq. (10) offers a new perspective on the powerful expressibility of the deep neural architectures.

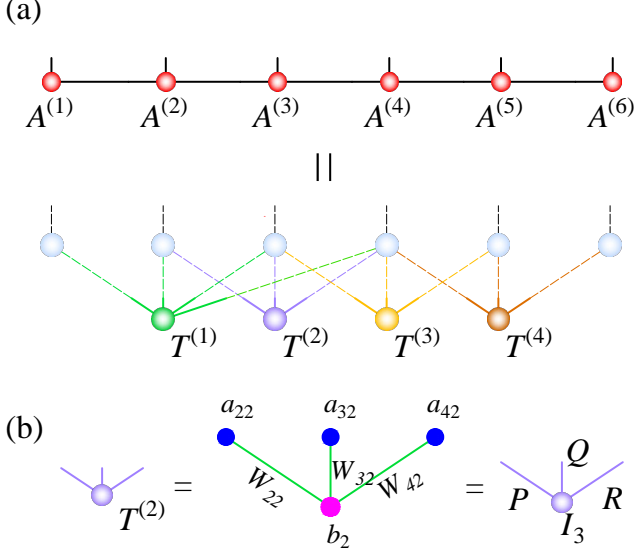


Figure 5. The sufficient and necessary conditions to transform an MPS to an RBM with a given architecture. (a) The RBM factorizes into n_h isolated tensors $T^{(1,2,3,4)}$ for given visible variables, see Eq. (12). (b) Decomposition of the tensor $T^{(2)}$ Eq. (13) can be carried out via the tensor rank decomposition Eq. (14).

III. RBM REPRESENTATION OF TNS: SUFFICIENT AND NECESSARY CONDITIONS

We now address the reverse question about how to transform a TNS to an RBM with a given architecture. It is meaningful to seek an RBM representation with given architectures because otherwise one can reproduce any function using an RBM with exponentially large resources [9–11]. We present a constructive approach to determine the RBM parameters from a TNS representation. Appendix B contains a sufficient condition with examples including the statistical Ising model and the cluster state [41].

To make the following discussions concrete, we present the general procedure using the MPS in Fig. 1(b) as an example. Assuming we seek the RBM parameters $\{W_{ij}, a_i, b_j\}$ of the RBM architecture shown in Fig. 1(a), we first require the MPS can be transformed into a TNS of structure shown in Fig. 5(a). The TNS factorizes into $n_h = 4$ isolated pieces with given physical indices. It is a product of the elements of one $2 \times 2 \times 2 \times 2$ tensor $T^{(1)}$ and three $2 \times 2 \times 2$ tensors $T^{(2,3,4)}$. Requiring the product equal to the MPS, we have

$$\text{Tr} \left(\prod_{i=1}^{n_v} A^{(i)} [v_i] \right) = T_{v_1 v_2 v_3 v_4}^{(1)} T_{v_2 v_3 v_4}^{(2)} T_{v_3 v_4 v_5}^{(3)} T_{v_4 v_5 v_6}^{(4)}. \quad (12)$$

Taking the logarithm of the above equation, we obtain $2^6 = 64$ linear equations for the $2^4 + 2^3 + 2^3 + 2^3 = 40$ tensor elements. These equations are therefore over determined. It is impossible to transform the MPS into the

RBM of the given structure in Fig. 1 suppose Eq. (12) has no solution. In that case, one can increase the number of variables by enlarging the hidden units or including more connectivities in the RBM. The linear equations will eventually become underdetermined with more complex RBM.

If a valid solution of Eq. (12) exists, we next aim to decompose each tensor $T^{(j)}$ into the components defined in Eqs. (4,5,6). See Fig. 5(b) for an example,

$$T_{v_2 v_3 v_4}^{(2)} = \sum_{h_2 \in \{0,1\}} e^{h_2 b_2 + \sum_{i \in \{2,3,4\}} v_i (W_{i2} h_2 + a_{i2})}, \quad (13)$$

where a_{ij} are partial contributions to the biases of the visible units. The overall biases sum up contributions from all the connections $a_i = \sum_j a_{ij}$. For a 3-index tensor $T^{(2)}$, there are in total $2^3 = 8$ equations with respect to 7 variables up to a normalization factor in Fig. 5(b). The number of variables grows linearly with the order of T , while the number of equations grows exponentially instead. The Eq. (13) is in general over determined and the solution exists only in special cases.

In practice, the Eq. (13) can be considered as the tensor rank decomposition (CP decomposition) [37, 38] of $T^{(2)}$,

$$T_{v_2 v_3 v_4}^{(2)} = \sum_{h_2 \in \{0,1\}} P_{v_2 h_2} Q_{v_3 h_2} R_{v_4 h_2}. \quad (14)$$

Note that in accordance to the RBM with the binary hidden units, the rank of the tensor decomposition should be two, i.e. P, Q, R are all 2×2 matrices.² For 3-index $2 \times 2 \times 2$ tensor, the decomposition of rank 2 is always possible by the Kronecker-Weierstrass normal form of the linear matrix pencil that the tensor represents [42]. For an arbitrary tensor, it is difficult to find the rank of tensor [37]. But it is much easier to judge whether it can be decomposed within a given rank. The rank is not smaller than the dimension of core tensor in higher order singular value decomposition[43]. If the dimension of the core tensor is larger than 2, the binary condition can not be satisfied. After the decomposition, we split each of these matrices into product of three matrices according to Eqs. (4-6) for the weight and biases of the RBM, for example

$$P = \begin{pmatrix} p & q \\ r & s \end{pmatrix} = p \begin{pmatrix} 1 & r/p \\ r/p & s/p \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & ps/qr \end{pmatrix} \begin{pmatrix} 1 & q/p \\ q/p & s/p \end{pmatrix}. \quad (15)$$

Comparing to Eqs. (4-6) we obtain

$$W_{ij} = \ln \frac{ps}{qr}, \quad (16)$$

$$a_{ij} = \ln \frac{r}{p}, \quad (17)$$

$$b_{ij} = \ln \frac{q}{p}, \quad (18)$$

² In general, if the rank of tensor decomposition is larger than two, one obtains an RBM with multistate hidden units.

where W_{ij} is the connection weight, a_{ij}, b_{ij} are partial contributions to the biases at the two ends of the connection. The overall biases sum up contributions from all the connections $a_i = \sum_j a_{ij}$ and $b_j = \sum_i b_{ij}$. In this way, each tensor of Eq. (12) is written in the RBM form, for example

Eqs. (12,13) are the sufficient and necessary conditions to translate a TNS into an RBM of a given architecture. In the case of $n_h = 1$, the Eq. (12) is just rephrases the MPS as the wavefunction tensor. It is however in general impossible to satisfy the subsequent rank two factorization requirement Eq. (13) or Eq. (14). By having more hidden units and connections one increases the degree of freedoms in Eqs. (12) and (13) until they both have solutions, which is in line with the mathematical results suggesting that the RBM can represent any function given sufficient large hidden units and dense connections [9–11].

In practice, a convenient way to quickly check whether a state has a particular RBM representation is to consider the factorization property Eq. (8), i.e. the RBM factorizes into a product state if one fixes a sequence of long enough visible units. For example, this necessary condition tells us that the Affleck-Kennedy-Lieb-Tasaki state [44] can not be written as an RBM (of ternary visible unit) with short range connections because of the existence of the hidden string order.

Finally, it is natural to map a TNS with translational invariance to an RBM with a uniform structure. For this purpose one can assemble tensors on a finite size torus and seek for the solutions of Eqs. (12,13).

IV. EXAMPLE: THE TORIC CODE STATES

As a concrete example of the discussed RBM-TNS correspondence, we present the RBM representation of the toric code states [45]. More examples including the statistical Ising model and the cluster state [41] are presented in the Appendix B.

The toric code model [45] has caught much attention of condensed matter and quantum information physicists because it is one of the simplest models that has possesses of topological order [46] and holds the promise for quantum information processing [47]. The toric code Hamiltonian reads

$$\mathcal{H} = - \sum_+ \mathcal{A}_+ - \sum_{\square} \mathcal{B}_{\square}, \quad (19)$$

$$\mathcal{A}_+ = \prod_{i \in +} \sigma_i^z, \quad \mathcal{B}_{\square} = \prod_{i \in \square} \sigma_i^x, \quad (20)$$

where Pauli matrices σ_i^x, σ_i^z reside on the bonds of the square lattice with periodic boundary condition. \mathcal{A}_+ consists of the product of four σ_i^z operators connecting to each vertex denoted by $+$, \mathcal{B}_{\square} consists of the product of four σ_i^x operators on each square plaquette \square . All the \mathcal{A}_+ and \mathcal{B}_{\square} terms commute.

The toric code ground state on a torus has four topologically sectors. Reference [18] already gave the RBM

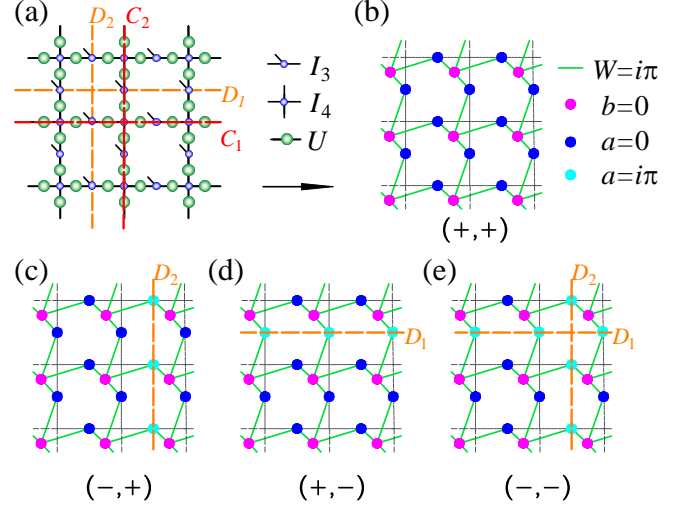


Figure 6. (a) The TNS and (b) the RBM representations of the toric code states. The meaning of each symbol is given in the figure. The dashed red and orange lines are the paths of Wilson loops defined in Eq. (26) and Eq. (27). (c-e) The RBM representations of the toric code states in the other topological sectors. The topological sectors are labelled by the eigenvalues of the Wilson loop operators (X_1, X_2) defined in Eq. (26). The cyan dots denote the visible units with bias $a = i\pi$.

representation of one of them via solving equations for the wavefunction coefficients. Now, given the RBM-TNS correspondence, we can readily construct the RBM representation of the toric code wavefunction in all of the four topological sectors based on their elegant TNS representations [22, 48].

We start with the bond dimension $D = 2$ PEPS representation of the toric code ground state [22, 48] shown in Fig. 6. The 4 index identity tensor I_4 is defined in vertices. The 3 index identity tensor I_3 resides on edges of lattice. $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ resides on the bond connecting I_4 and I_3 . We identify the bond centered I_3 tensors as the visible units and the tensors I_4 on the vertices as the hidden units of the RBM. Following the recipes of Eqs. (16-18), we decompose the matrix

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \\ & e^0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & e^{i\pi} \end{pmatrix} \begin{pmatrix} 1 & \\ & e^0 \end{pmatrix}, \quad (21)$$

corresponding to the weights and biases of the RBM,

$$W = i\pi, \quad (22)$$

$$a = 0, \quad (23)$$

$$b = 0. \quad (24)$$

The resulting RBM representation of the toric state is shown in Fig. 6(b). The magenta hidden unit couples to the four visible variables around each vertex. Tracing out the hidden units, the RBM function defined in Eq. (2) is

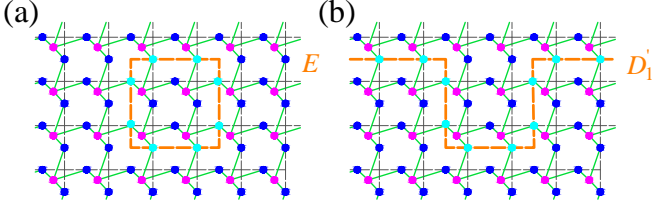


Figure 7. The symbols are the same as Fig. 6. (a) Acting four \mathcal{A}_+ operators [Eq. (20)] in the center changes the visible biases along the path E to $i\pi$. The RBM function is still in the $(+, +)$ sector identical to Fig. 6(b). (c) The RBM wavefunction remains in the $(+, -)$ sector as Fig. 6(d) after four \mathcal{A}_+ operations. The operation moves the path D_1 in Fig. 6(d) to D_1' .

a product over all the vertices

$$\Psi_{\text{TC}}(v) = \prod_{+} \left(1 + e^{i\pi \sum_{i \in +} v_i}\right), \quad (25)$$

which requires an even number of visible units with $v_i = 1$ at each vertex. The RBM state thus represents an equal weight superposition of the closed loop configurations as the toric ground state [45]. This construction is simpler than Ref. [18] which introduced hidden units on the vertices as well as in the plaquette centers.

The topological property of the toric code state can also be demonstrated in the RBM language. We first define two Wilson loop operators

$$X_1 = \prod_{i \in C_1} \sigma_i^x, \quad X_2 = \prod_{i \in C_2} \sigma_i^x, \quad (26)$$

where C_1, C_2 denote the paths along the x/y directions indicated by the red lines in Fig. 6(a). The Wilson loop operators X_1, X_2 are mutually commuting and they commute with the Hamiltonian (19). So we can label the toric code ground states using the eigenvalues of Eq. (26). There are totally four topological sectors. The wavefunction Eq. (25) illustrated in Fig. 6(b) belongs to the $(+, +)$ sector, we denote it by $\langle \{v_i\} | \Psi^{(+,+)} \rangle = \Psi_{\text{TC}}(\{v_i\})$.

Next, we define another two operators along the paths D_1, D_2 indicated by the dashed orange lines in Fig. 6(a),

$$Z_1 = \prod_{i \in D_1} \sigma_i^z, \quad Z_2 = \prod_{i \in D_2} \sigma_i^z, \quad (27)$$

These two operators do not commute with X_1, X_2 . They translate the states into different topological sectors [49]. $|\Psi^{(-,+)}\rangle = Z_2 |\Psi^{(+,+)}\rangle$, $|\Psi^{(+,-)}\rangle = Z_1 |\Psi^{(+,+)}\rangle$, $|\Psi^{(-,-)}\rangle = Z_1 Z_2 |\Psi^{(+,+)}\rangle$. In the PEPS representation, the operators Z_1, Z_2 modify the identity tensor I_3 in Fig. 6(a) into a diagonal tensor $\text{diag}(1, -1)$. Accordingly, this changes the bias of the visible units in Eq. (23) to $a = \ln(-1) = i\pi$ along the D_1 and D_2 lines [Fig. 6(c,d,e)].

Note that the Wilson lines $D_{1,2}$ do not need to be straight lines. It is only crucial that they wind around the torus, see Fig. 7. The X_1 defined along D_1 [Fig. 6(d)]

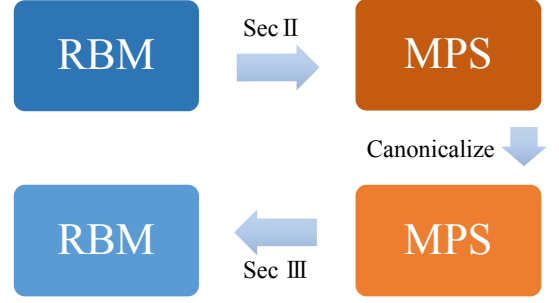


Figure 8. Simplify an RBM by mapping it into an MPS and then transforming the canonicalized MPS back. For detailed discussions of the mappings, see Sec. II and Sec. III.

and D_2' are related by a loop of σ_i^z operators along E in Fig. 7(b). It is the product of the \mathcal{A}_+ operator surrounded by the path E . Since the ground state is the eigenstate of every term in the toric code Hamiltonian Eq. (19), acting \mathcal{A}_+ to arbitrary vertex will not change the wavefunction. In Fig. 7(a), the visible bias along the closed loop E is $a = i\pi$. The wavefunction is the same as Fig. 6(b) because the path does not winds around the torus. Similarly, the state in Fig. 7(b) is identical to Fig. 6(c). The redundancy of the RBM parametrization will be discussed in detail in Sec. V.

The PEPS wavefunctions can be decomposed into products of translationally invariant tensors in the form of Eq. (12) in Sec. III. Conversely, starting from the RBM representation Eq. (25) we will obtain the PEPS back following the instructions of Sec. II.

V. SIMPLIFYING THE RBM USING TNS METHODS

Similar to TNS, it is known that there are redundancy in the neural network functions [3]. A network with different connection weights and biases may result into the same function. The knowledge about the redundancy in TNS can help simplify the RBMs.

Figure 8 illustrated the idea with diagrams. Given an RBM, we first obtain its MPS representation following the algorithms of Sec. II. We then transform the MPS into its canonical form by performing the Schmidt decomposition at each bond [34, 35]. This will reduce the bond dimensions to the minimal requirement and fix the gauge degree of freedoms. Finally, we map the canonicalized MPS to another RBM following the procedure of Sec. III. Because the canonicalized MPS has smaller bond dimensions, the final RBM will be simpler than the original one.

As an example of the above procedure, we consider the RBM representation of the 1D cluster state given in Ref. [18]. The RBM state has $n_h = n_v$ hidden units with

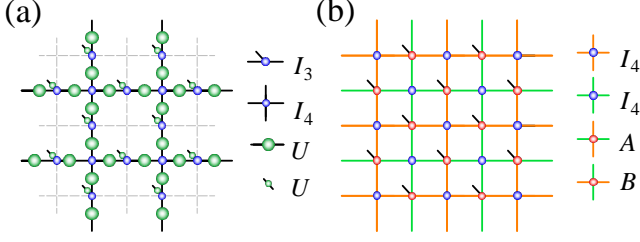


Figure 9. Equivalent TNS representations of the toric code ground states. (a) The simplex is defined in the plaquette. The green dot is the matrix $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. (b) The simplex is defined both in vertex and plaquette. The 5 index tensor A, B are given by projectors acting on an arbitrary state.

each hidden unit connecting to three visible units. According to Fig. 2 it maps to a $D = 4$ MPS because each bipartition will cut $n = 2$ connections. However, canonicalization of the MPS or using the scheme of Fig. 4 will reduce the bond dimension to $D = 2$. Mapping the simplified MPS back, we obtain an RBM with each hidden unit connecting only to two neighboring visible units. See Appendix B for detailed discussions of this RBM architecture and the explicit parameters in Tab. I. MPS representation of the RBM function, and in particular its canonical form, is very useful to reveal the essence of the state dictated by the entanglement entropy requirement.

The situation is however more complicated if one maps an RBM to a higher dimensional TNS such as PEPS. There is no general numerical approach to remove the redundancy in the PEPS. There are two cases,

1) The trivial case, where the two PEPSs can be transformed into each other by gauge transform in the bonds. This corresponds to the different assignment of the cross gate in Fig. 2.

2) The non-trivial case. The two PEPSs are so different that we can't transform one into the other with only gates acting on the bonds. This is due to the non-injectivity property [50] of the wavefunction. Take the toric code state in Sec. IV as an example. We can act the projection operator $\mathcal{P} = \prod_{+} \frac{1+A_{+}}{2} \prod_{\square} \frac{1+B_{\square}}{2}$ to arbitrary wavefunctions to obtain the ground state. The construction is therefore not unique and the TNS of toric code is not injective [51] either. The state shown in Fig. 6(a) is a result of projection on the state $\left(\frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}} \right)^{\otimes n_v}$. If we instead project onto the state $\mathcal{P} |\uparrow\rangle^{\otimes n_v}$ we will obtain another PEPS with the 4 index tensor defined in the plaquette, see Fig. 9(a). More generally, we can apply the projector \mathcal{P} to an arbitrary product state to obtain the toric code ground state. For example, the TNS can have 4 index tensors defined both at the vertices and at plaquette center as shown in Fig. 9(b). The corresponding RBM representation is the one discussed in the Ref. [18]. Although they are all equivalent states, it is impossible to find a local transformation to connect the local tensor be-

cause the ground state is a Z_2 spin liquid state [52, 53]. Since the redundancies in non-injective case are richer, it offers us more possibility of simplifying the RBM via different TNS constructions.

VI. SUMMARY AND DISCUSSION

The constructive connection between RBM and TNS functions bridges between the worlds of deep learning and quantum many-body physics. The RBM-TNS correspondence has broad implications for both communities. The entanglement entropy bound of the TNS can rigorously quantify the expressive power of the RBM functions. This physics perspective provides tighter bounds on the number of hidden units and the connectivities of the RBMs than previously known results solely based on mathematical considerations [9–11]. Akin to the success of the TNS methods in the quantum physics, our finding suggests that the success of RBM learning may be related to the relatively low entanglement entropy in the typical datasets fulfilling the physics laws such as the natural images and speech signals. Exploiting the RBM-TNS correspondence one may put the previously argued operational similarity between the renormalization group and deep learning [7] to a more rigorous and constructive manner. On the technical side, TNS methods have already been used for pattern recognition [54]. Given the deeper connection between the two fields suggested by the RBM-TNS correspondence, more insights and computational tools of both fields can be exchanged. For example, the canonical MPS [34, 35] and the tensor renormalization group methods [55] can be employed to transform and simplify the RBM architectures.

On the other hand, the RBM is a more compact variational wavefunctions [17] because it can represent the TNS with much fewer parameters, see discussions in Sec. II. Deep learning algorithms and industrial software/hardware [56–58] may thus be beneficial for quantum many-body physics researches through this connection.

For future research directions, it is interesting to further explore the connection between deep learning architectures such as the deep Boltzmann machines [31] and the TNS with hierarchical structures such as tree tensor networks [59–62] and multi-scale entanglement renormalization ansatz [39]. In passing we also note the efforts of understanding the expressive power of the deep feedforward neural networks [63, 64]. We believe that the insights of the quantum entanglement and tensor networks can provide even deeper understanding of the deep learning and guides better neural nets design.

VII. ACKNOWLEDGMENTS

We thank Dongling Deng, Xiaopeng Li, Chen Fang, Hong-Hao Tu and Yi-feng Yang for helpful discussions.

This work is supported by the National Natural Science Foundation of China (Grants No. 11474331 and No.

11190024) and the Ministry of Science and Technology of China (Grant No. 2016YFA0302400).

-
- [1] Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521**, 436 (2015).
 - [2] I. Goodfellow, Y. Bengio, and A. Courville, “Deep learning,” (2016), book in preparation for MIT Press.
 - [3] C. M. Bishop, *Neural Networks for Pattern Recognition*, 1st ed. (Clarendon Press, Oxford : New York, 1996).
 - [4] S. S. Haykin, *Neural networks and learning machines*, Vol. 3 (Pearson Upper Saddle River, NJ, USA, 2009).
 - [5] M. A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, 2015).
 - [6] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **32**, 1007 (1985).
 - [7] P. Mehta and D. J. Schwab, ArXiv e-prints (2014), [arXiv:1410.3831 \[stat.ML\]](https://arxiv.org/abs/1410.3831).
 - [8] H. W. Lin and M. Tegmark, ArXiv e-prints (2016), [arXiv:1608.08225 \[cond-mat.dis-nn\]](https://arxiv.org/abs/1608.08225).
 - [9] Y. Freund and D. Haussler, “Unsupervised learning of distributions of binary vectors using two layer networks,” (1994).
 - [10] N. Le Roux and Y. Bengio, *Neural computation* **20**, 1631 (2008).
 - [11] G. Montufar and N. Ay, *Neural computation* **23**, 1306 (2011).
 - [12] G. E. Hinton, *Neural computation* **14**, 1771 (2002).
 - [13] T. Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient* (ACM, New York, New York, USA, 2008).
 - [14] <http://www.deeplearning.net/tutorial/rbm.html>.
 - [15] G. E. Hinton and R. R. Salakhutdinov, *Science* **313**, 504 (2006).
 - [16] G. E. Hinton, S. Osindero, and Y. W. Teh, *Neural computation* **18**, 1527 (2006).
 - [17] G. Carleo and M. Troyer, ArXiv e-prints (2016), [arXiv:1606.02318 \[cond-mat.dis-nn\]](https://arxiv.org/abs/1606.02318).
 - [18] D.-L. Deng, X. Li, and S. Das Sarma, ArXiv eprint (2016), [arXiv:1609.09060 \[cond-mat.dis-nn\]](https://arxiv.org/abs/1609.09060).
 - [19] G. Torlai and R. G. Melko, *Physical Review B* **94**, 165134 (2016).
 - [20] L. Huang and L. Wang, *Phys. Rev. B* **95**, 035105 (2017).
 - [21] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, *Phys. Rev. B* **95**, 041101 (2017).
 - [22] R. Orús, *Annals of Physics* **349**, 117 (2014).
 - [23] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, eprint [arXiv:quant-ph/0608197](https://arxiv.org/abs/quant-ph/0608197) (2006), [quant-ph/0608197](https://arxiv.org/abs/quant-ph/0608197).
 - [24] I. V. Oseledets, *SIAM Journal on Scientific Computing* **33**, 2295 (2011).
 - [25] F. Verstraete and J. I. Cirac, eprint [arXiv:cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066) (2004), [cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066).
 - [26] F. Verstraete, V. Murg, and J. I. Cirac, *Advances in Physics* **57**, 143 (2008).
 - [27] J. Eisert, M. Cramer, and M. B. Plenio, *Rev. Mod. Phys.* **82**, 277 (2010).
 - [28] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
 - [29] E. Stoudenmire and S. R. White, *Annual Review of Condensed Matter Physics* **3**, 111 (2012).
 - [30] G. E. Hinton and T. J. Sejnowski, in *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1 (1986) pp. 282–317.
 - [31] R. Salakhutdinov and G. E. Hinton, *JMLR Workshop Conf. Proc.* **5**, 448 (2009).
 - [32] See <https://github.com/yzcj105/rbm2mps> for the Matlab implementations of the RBM to MPS transformation corresponding to Fig. 2 and Fig. 4 respectively.
 - [33] B. M. Marlin, K. Swersky, B. Chen, and N. de Freitas, *JMLR Workshop Conf. Proc.* **9**, 509 (2010).
 - [34] G. Vidal, *Phys. Rev. Lett.* **91**, 147902 (2003).
 - [35] G. Vidal, *Phys. Rev. Lett.* **93**, 040502 (2004).
 - [36] U. Schollwöck, *Annals of Physics* **326**, 96 (2011), [arXiv:1008.3477 \[cond-mat.str-el\]](https://arxiv.org/abs/1008.3477).
 - [37] https://en.wikipedia.org/wiki/Tensor_rank_decomposition.
 - [38] F. L. Hitchcock, *J. Math. Phys* **6**, 164 (1927).
 - [39] G. Vidal, *Phys. Rev. Lett.* **99**, 220405 (2007).
 - [40] G. Evenbly and G. Vidal, *Physical review letters* **112**, 240502 (2014).
 - [41] H. J. Briegel and R. Raussendorf, *Phys. Rev. Lett.* **86**, 910 (2001).
 - [42] J. M. Landsberg, *Tensors: geometry and applications*, Vol. 128 (American Mathematical Society Providence, RI, USA, 2012).
 - [43] T. G. Kolda and B. W. Bader, *SIAM review* **51**, 455 (2009).
 - [44] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, *Phys. Rev. Lett.* **59**, 799 (1987).
 - [45] A. Kitaev, *Annals of Physics* **303**, 2 (2003).
 - [46] X. G. Wen, *Quantum Field Theory of Many-Body Systems* (Oxford University Press, 2007).
 - [47] S. B. Bravyi and A. Y. Kitaev, eprint [arXiv:quant-ph/9811052](https://arxiv.org/abs/quant-ph/9811052) (1998), [quant-ph/9811052](https://arxiv.org/abs/quant-ph/9811052).
 - [48] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, *Phys. Rev. Lett.* **96**, 220601 (2006).
 - [49] A. Kitaev, A. Shen, and M. Vyalii, *Classical and Quantum Computation*, Graduate studies in mathematics (American Mathematical Society, 2002).
 - [50] N. Schuch, I. Cirac, and D. Pšřez-García, *Annals of Physics* **325**, 2153 (2010).
 - [51] N. Schuch, D. Poilblanc, J. I. Cirac, and D. Pérez-García, *Phys. Rev. B* **86**, 115108 (2012).
 - [52] N. Read and S. Sachdev, *Phys. Rev. Lett.* **66**, 1773 (1991).
 - [53] X. G. Wen, *Phys. Rev. B* **44**, 2664 (1991).
 - [54] E. Miles Stoudenmire and D. J. Schwab, ArXiv e-prints (2016), [arXiv:1605.05775 \[stat.ML\]](https://arxiv.org/abs/1605.05775).
 - [55] M. Levin and C. P. Nave, *Phys. Rev. Lett.* **99**, 120601 (2007).
 - [56] <https://www.nvidia.com/en-us/deep-learning-ai/developer>.
 - [57] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G.

- Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, ArXiv e-prints (2016), [arXiv:1605.08695 \[cs.DC\]](#).
- [58] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, ArXiv e-prints (2016), [arXiv:1603.04467 \[cs.DC\]](#).
 - [59] V. Murg, F. Verstraete, O. Legeza, and R. M. Noack, *Phys. Rev. B* **82**, 205105 (2010).
 - [60] P. Silvi, V. Giovannetti, S. Montangero, M. Rizzi, J. I. Cirac, and R. Fazio, *Phys. Rev. A* **81**, 062335 (2010).
 - [61] Y.-Y. Shi, L.-M. Duan, and G. Vidal, *Phys. Rev. A* **74**, 022320 (2006).
 - [62] L. Tagliacozzo, G. Evenbly, and G. Vidal, *Phys. Rev. B* **80**, 235127 (2009).
 - [63] N. Cohen, O. Sharir, and A. Shashua, *JMLR* (2015).
 - [64] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, and Q. Liao, ArXiv e-prints (2016), [arXiv:1611.00740 \[cs.LG\]](#).
 - [65] K. Sohn and H. Lee, ArXiv e-prints (2012), [arXiv:1206.6418 \[cs.LG\]](#).
 - [66] J. Hubbard, *Physical Review Letters* **3**, 77 (1959).

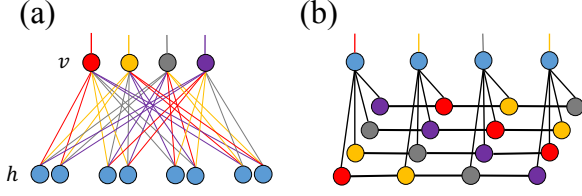


Figure 10. (a) The shift-invariant RBM. The number of hidden units are enlarged by a factor of n_v because of the permutation of the connections. The state respects translation invariance of the visible variables by construction. (b) The corresponding MPS is constructed by connecting n_v copies of the same MPS, each tensor offsets by one site.

Appendix A: TNS representation of the Shift-Invariant RBM and its entanglement bound

The variational Monte Carlo study [17] employed a shift-invariant RBM function [65] to respect the translational invariance of the physical systems. The variational ansatz is a product of n_v RBM functions defined in Eq. (2)

$$\Psi(v) = \prod_{\mathcal{T}} \Psi_{\text{RBM}}(\mathcal{T}\{v_i\}), \quad (\text{A1})$$

where \mathcal{T} is the translational operator which shifts the visible variables to equivalent positions. Fig. 10(a) shows the shift-invariant product Eq. (A1) as a single RBM. Assuming each factor of Eq. (A1) contains n_h hidden units, the enlarged RBM has $n_v n_h$ hidden variables.

This shift-invariant RBM can be written as a single MPS in two ways. Naively, one can first rewrite each factor of Eq. (A1) as an MPS with bond dimension D , then multiply these MPSs together according to Fig. 10(b). The resulting MPS has bond dimension D^{n_v} , which appears to have a volume law entanglement entropy bound even for constant D . However, the tighter bound Eq. (10) of Sec. II shows that multiplying the RBM according to Eq. (A1) does not increase the entanglement entropy bound compared to a single RBM, because the effective connectivity between the visible units does not increase with the enlarged visible-hidden connections, see Fig. 10(a).

The connection of the shift-invariant RBM to MPS justifies the excellent accuracy reached in the variational study [17]. On the other hand, the correspondence also shows that to have the entanglement capacity beyond the standard tensor network states one has to use dense connections in each factor of Eq. (A1).

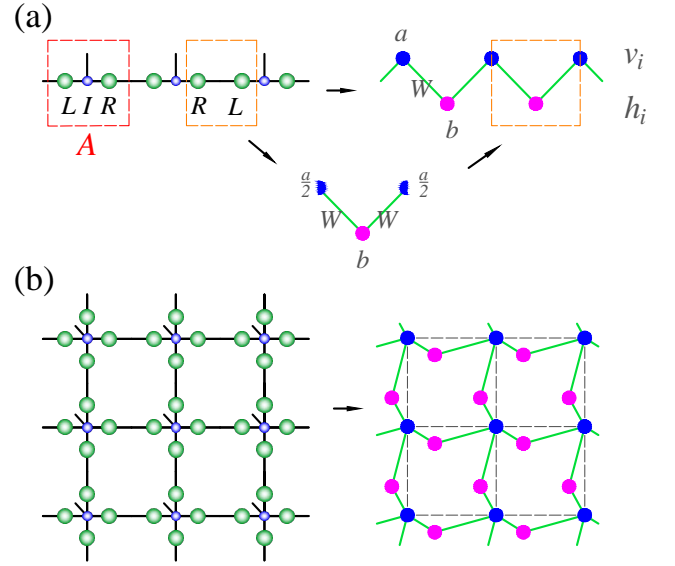


Figure 11. (a) The tensor A in the dashed red box has the special form Eq. (B1). The green dots denote the L and R matrices, while the blue dots are the 3 index identity tensor. We transform the matrix product in the dashed orange rectangular according to Eq. (B2) to obtain the RBM parameters. The half blue dot denotes $a/2$ since the visible unit is shared by two neighboring dashed orange boxes. (b) The PEPS of a similar structure is mapped to an RBM. The meaning of the symbols is the same as (a).

Appendix B: A sufficient condition for RBM representation of MPS/PEPS and examples

We give a sufficient condition for the MPS or PEPS to have an RBM representation. Many physically interesting thermal states and quantum wavefunction belong to this class. For example, the toric code model discussed in Sec. IV, the statistical Ising model with external field, and the 1D/2D cluster states discussed in this section.

A sufficient condition for an MPS to have the RBM representation is that each tensor has the following form

$$A_{\alpha\beta}[v] = L_{\alpha v} R_{v\beta}, \quad (\text{B1})$$

where L and R are two 2 by 2 matrices. As shown in Fig. 11(a), the product of R and L in orange dashed box can be replaced by coupling to the hidden unit of the RBM. The bias for the visible unit (blue dot) is $a/2$ because it is shared by two neighboring boxes. According to Eqs. (4-6), we can write the RL into the RBM parameters

$$RL = \begin{pmatrix} 1 & e^{a/2} \\ 1 & e^W \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & e^W \end{pmatrix} \begin{pmatrix} 1 & e^b \\ 1 & e^W \end{pmatrix} \quad (\text{B2})$$

The decomposition can be arbitrary. Here we choose a symmetric form for simplicity.

One example of this type is the statistical Ising model with the partition function,

$$Z = \sum_{\{s_i\}} \exp \left(K \sum_{\langle i,j \rangle} s_i s_j + H \sum_i s_i \right) \quad (\text{B3})$$

where K is the coupling constant and H is the external field. $s_i = \pm 1$ are the Ising spins. To rewrite the partition function into a summation of the RBM function of the form Eq. (2) we introduce binary variables $v_i = (s_i + 1)/2$.

In one dimension, the Ising partition function can be represented as an MPS shown in Fig. 11. The matrix product on each bond reads

$$RL = \begin{pmatrix} e^{K+H} & e^{-K} \\ e^{-K} & e^{K-H} \end{pmatrix}. \quad (\text{B4})$$

Combining Eq. (B2) and Eq. (B4), we obtain the RBM parameters $\{W, a, b\}$ summarized in the first line of Tab. I.

This procedure can be readily generalized to 2D, where the partition function is represented by a PEPS. The PEPS tensor follows a condition similar to Eq. (B1), as illustrated in Fig. 11(b). We introduce one hidden unit for each bond which couples to the two visible units connected by the lattice bond. The only difference compared to the 1D case is that the one replaces H by $H/2$ in the RL matrix Eq. (B4) since the each site is shared by 4 instead of 2 bonds. Correspondingly, the $a/2$ bias in Eq. (B2) is replaced by $a/4$ because the visible bias is also shared by 4 connections [Fig. 11(b)]. The RBM parameters for 2D Ising model are summarized in the second line of Tab. I.

The above results show that a very simple sparse RBM with $n_h = n_v$ (or $n_h = 2n_v$ in 2D) hidden units defined on the bonds can exactly reproduce the thermal distribution of the Ising model. It is remarkable that this is independent of the coupling strength and holds even at the criticality where the correlation between the visible spins are long ranged [19]. Essentially, the effect of the hidden units of the RBM play the role of a Hubbard-Stratonovich auxiliary field which decouple the interaction on the bond [66].

Another example of the TNS satisfying Eq. (B1) is the cluster state [41]. The MPS representation [22, 48] has $RL = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ on every bond. We can obtain the RBM parameters W, a, b using Eq. (B2) summarized in the Tab. I. Similar to the Ising model case, the RBM has hidden units coupled to the physical degree of freedoms on each lattice bond. These construction is simpler than the one of Ref. [18] which requires each visible unit to connect to 3 hidden units. The simplification is due to we construct the RBM representation from the canonical MPS of the cluster state.

Appendix C: General equivalence between Boltzmann machines and TNS

The name ‘‘Restricted’’ in the RBM means that there are only connections between the visible and hidden units, not within themselves. Our results on RBM-TNS correspondence can also be generalized to the case without such restrictions. For example, the deep Boltzmann machines [31] have multilayers of hidden units with interconnections, and the Boltzmann machines (BM) [30] have direct connections within the visible and hidden units.

In general, the BM parametrizes a function in the form

$$\Psi_{\text{BM}}(v) = \sum_h e^{-E(v,h)}, \quad (\text{C1})$$

with the energy function

$$E(x = v \cup h) = - \left(\sum_{i,j} W_{ij} x_i x_j + \sum_k \theta_k x_k \right), \quad (\text{C2})$$

where W_{ij} is the connection weight between the units i and j , and θ_k is the bias of the unit k . One can either view Eq. (C1) as a probability distribution or a complex wavefunction amplitude.

To write Eq. (C1) into a TNS, we introduce tensor $M^{(ij)} = \begin{pmatrix} 1 & 1 \\ 1 & e^{W_{ij}} \end{pmatrix}$ on the edges and diagonal tensors $T^{(k)} = \text{diag}(1, e^{\theta_k})$ on the vertices. For the visible units, the diagonal tensors have an additional dimension corresponding to the external degree of freedoms. Using these tensors, the BM Eq. (C1) can be written as a tensor network state

$$\Psi_{\text{TNS}}(v) = \text{Tr} \left(\prod_{i,j} M^{(ij)} \prod_k T^{(k)} \right). \quad (\text{C3})$$

Conversely, one can also attempt to map a TNS back to a BM. First, we prove that any TNS with only rank 2 tensors can be directly mapped back to a binary BM. The CP decomposition or rank decomposition [37] of $d_1 \times d_2 \times \dots \times d_n$ tensor T is given by

$$T_{i_1 i_2 \dots i_n} = \sum_{k=1}^r \lambda_k P_{i_1 k}^{(1)} P_{i_2 k}^{(2)} \dots P_{i_n k}^{(n)}, \quad (\text{C4})$$

where $P^{(1)}, P^{(2)}, \dots, P^{(n)}$ are matrix. There exist a finite minimum r so that the equation holds. If T is a vector, $r = 1$. If T is a matrix, r equals the small dimension of the matrix. However, when $n \geq 3$, r is not necessarily small than any d_i . r is the rank of a tensor. That is to say any tensor can be written as a contract of a diagonal tensor constructed by $\lambda_k (k = 1, 2, \dots, r)$ and n matrix $P^{(i)} (i = 1, 2, \dots, n)$ with dimension $d_i \times r$. If $d_i = 2, \forall i$ and $r = 2$, it can be mapped to a binary BM. When contracting the α -th index of $T^{(i)}$ and β -th index of $T^{(j)}$,

Model	W	b	a
1D Ising	$\ln(4e^{4K} - 2)$	$-\ln(e^{4K} - 1) - \ln 4$	$-4K - 2H - 2\ln 2$
2D Ising	$\ln(4e^{4K} - 2)$	$-\ln(e^{4K} - 1) - \ln 4$	$-8K - 2H - 4\ln 2$
1D cluster	$\ln \frac{3}{2}$	$-\ln 2 + i\pi$	$2\ln 2$
2D cluster	$\ln \frac{3}{2}$	$-\ln 2 + i\pi$	$4\ln 2$

Table I. The RBM parameters for the statistical Ising model and quantum wavefunction. Each hidden unit interacts with two visible units connected by a bond. The parameterizations are not unique and we only list one of them. The meaning of the Ising model parameters is given in Eq. (B3)

we can do rank decomposition and multiplies the matrix together.

$$M^{(ij)} = P^T Q = \begin{pmatrix} p & q \\ r & s \end{pmatrix}, \quad (\text{C5})$$

where P (Q) are the matrix decomposed by $T^{(i)}(T^{(j)})$ in Eq. (C4). Comparing with Eqs. (15,16), we find that

$W_{ij} = \ln \frac{ps}{qr}$. The two diagonal matrix in Eq. (15) can be absorbed into the diagonal tensor, which makes a contribution to θ_i . W_{ij} and θ_i need to be complex number if the base of logarithm is negative. The result BM shares the same configuration with TNS. For a general TNS with larger ranks, we can always perform the CP decomposition follow the same procedure. The resulting BM will have multistate units.